

Detecting Players Personality Behavior with Any Effort of Concealment

Fazel Keshtkar, Candice Burkett, Arthur Graesser, and Haiying Li

Institute for Intelligent Systems, The University of Memphis
Memphis, TN, USA

{fkshtkar,cburkett,a-graesse,hli15}@memphis.edu

Abstract. We introduce a novel natural language processing component using machine learning techniques for prediction of personality behaviors of players in a serious game, Land Science, where players act as interns in an urban planning firm and discuss in groups their ideas about urban planning and environmental science in written natural language. Our model learns vector space representations for various features extraction. In order to apply this framework, input excerpts must be classified into one of six possible personality classes. We applied this personality classification task using several machine learning algorithms, such as: Naïve Bayes, Support Vector Machines, and Decision Tree. Training is performed on a relatively dataset of manually annotated excerpts. By combining these features spaces from psychology and computational linguistics, we perform and evaluate our approaches to detecting personality, and eventually develop a classifier that is nearly 83% accurate on our dataset. Based on the feature analysis of our models, we add several theoretical contributions, including revealing a relationship between different personality behaviors in players' writing.

Keywords: Personality Detection, Classification, Conversation, Leary's Rose Framework, Natural Language Processing, Sentiment Analysis.

1 Introduction

Detecting personality and/or behavior in conversation is a hard task. In serious game (i.e., in chat rooms in serious games), players may have talk about different ideas than others they are chatting with during conversation. They also might be expose to or be affected with different personalities or moods during the conversation by other players. On the other hand, players have various personality behavior, such as, helping, leading, aggressive, or dependent. Their personalities may cause them to behave varied within conversation. In our model, we aim to detect player's personality preferably without disrupting their relationship with others. We believe these findings will help human mentors manage players and interact in the right manner in conversation.

In this paper, we explore a component of natural language processing, using machine learning techniques, based on Leary's Rose framework (See Section 1.2)

for prediction of personality behaviors of players in a serious game, Land Science, where players act as interns in an urban planning firm and discuss in groups their ideas about urban planning and environmental science in written natural language. The possible dialog paths are defined by Leary Rose, a framework for interpersonal communication.

Our model learns vector space representations for various feature extraction. In order to apply this framework, input excerpts must be classified into one of six possible personality classes (See Section 1.2). We applied this personality classification task using several machine learning algorithms. More specifically, classification performance was measured using a Naïve Bayes classifier, Support Vector Machines algorithm, and Decision Tree named J48. And the raining set is performed on a relatively dataset of manually annotated excerpts. We extract a combination of features from psychology and computational linguistics. We develop and evaluate our approaches to detecting personality, and eventually develop a classifier that is nearly 80% accurate on our dataset. Based on feature analysis of our models, we add several theoretical contributions, including a relationship between different personality behavior in players writing.

1.1 Land Science Game

Land Science is a “serious game” created by researchers at the University of Wisconsin-Madison [1,14,3] that has been designed to simulate a regional planning practicum experience for students. During the 10 hour game, students play the role of interns at a fictitious regional planning firm (called Regional Design) where they make land use decisions in order to meet the desires of virtual stakeholders who are represented by Non-Player Characters (NPCs). Students are split into groups and progress through a total of 15 stages of the game in which they complete a variety of activities including a virtual site visit of the community of interest in which students familiarize themselves with the history and ecology of the area as well as the desires of difference stakeholder group. In addition, students get feedback from the stakeholders, and use a custom designed Geographic Information System (iPlan) to create a regional design plan. Throughout the game players communicate with other members of their planning team as well as a mentor (i.e., an adult who is representing a professional planner with the fictitious planning firm) through the use of a chat feature that is embedded in the game.

1.2 Leary’s Rose Framework

Leary’s Interpersonal Circumplex (also referred to as Leary’s Rose) has been used by researchers for many decades as a foundation for categorizing personality characteristics based on the statements people make [9]. Leary’s circumplex measures characteristics on two dimensions: the above-below axis represents variations from dominant (above) to submissive (below) and the opposed-together axis represents variations of cooperation from accommodating (together) to rebellious (opposed). The use of two axes allows the Rose to be easily separated into

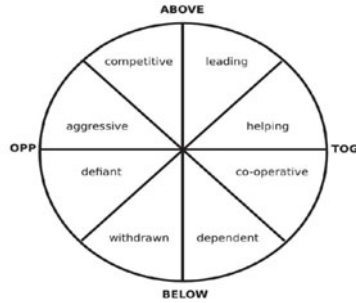


Fig. 1. Leary’s Rose Framework

Table 1. Our dataset with some examples of student’s conversations that convey Leary’s Rose categories

Category	Percent	Example
Leading	13.69%	Finish your task now so we can move on.
Helping	22.22%	How can I help you with that?
Competitive	24.48%	My plan is better than your plan.
Aggressive	03.04%	That idea is stupid. It will never work.
Dependent	29.98%	What should I do now?
Withdrawn	04.71%	Sorry, never mind, I’m not thinking.

4 quadrants: above-together, above-opposed, below-together and below-opposed. Furthermore, each of these quadrants can be further split resulting in a total of eight different characteristic indices (see Figure 1).

The rest of the paper is organized as in the followings. Our model is described in Section 2 as well as dataset and human annotation performance. The Section 3 provides a summary of the experiments and results along with discussion. In Section 4, we describe related works and addressing personality detection in similar context, e.g., online chats. Finally, Section 5 is the conclusion and addresses the future work direction.

2 Our Model

2.1 Dataset Construction

Players in the epistemic game, Land Science, communicate with both other players and mentors using chat windows embedded in the game. For the purposes of these analysis, we only assessed the discourse of the players and did not analyze the discourse of the mentors. Annotation was done using the coding scheme (further discussed under Human Annotation) that was developed by the researchers based on the Timothy Leary’s Interpersonal Circumplex Model [9]. The researchers selected 1,000 player excerpts (average length = 4.8 words) to

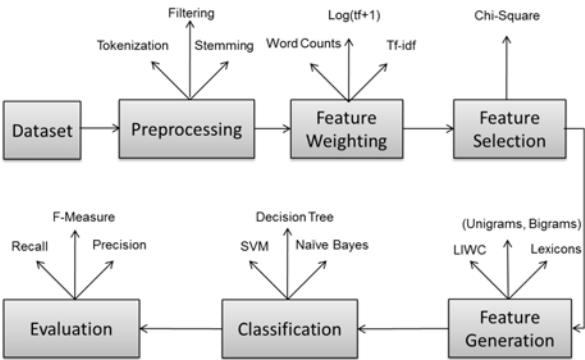


Fig. 2. Leary’s Rose Framework

be analyzed. For our purposes, an excerpt was defined as a turn of speech that was taken by the student. On the other word, in the one excerpt occurred every time a student typed something and clicked “send” or hit “enter” in the chat function. The excerpts were selected from a larger set of 3,227 excerpts, so approximately 31% of the excerpts were used in the analyzed data set. In order to proportionally represent all stages of the game in the set that was analyzed, approximately 31% of the player excerpts were randomly selected from each stage of the game. Our model is illustrated in Figure 2 and in the following sections we describe the components of this model.

Human Annotation. For the purposes of this study, researchers developed a coding scheme based on Leary’s Interpersonal Circumplex. This coding scheme looked specifically at all 4 quadrants of the Circumplex, but combined Helping and Co-operating into one category simply referred to as “Together” and also combined Aggressive and Defiant into once category simply referred to as “Opposed”. In other words, the developed coding scheme focused on 6 categories: Competitive, Leading, Dependent, Withdrawn, Opposed and Together. Using this coding scheme, two trained researchers annotated the data set of 1,000 excerpts. The first series of training required the human annotators to independently code 200 excerpts randomly selected from the Land Science corpus. The kappa statistic was computed to assess inter-rater reliability on this set and agreement was fair (.33). Following this, the annotators discussed and refined any issues regarding the coding scheme and then annotated a new set of 1,000 excerpts that were randomly selected from the Land Science corpus. The kappa statistic was computed to assess inter-rater reliability on the second training set and agreement was substantial (.70). Results indicated increased reliability and thus completed the training of the human annotators. Once the two annotators were trained they independently annotated a set of 1,000 excerpts described in the data set portion of this paper.

Lexicon Resources. Sentiment-based lexical resources annotate words/concepts with polarity. To achieve greater coverage, we use four different sentiment-based lexical resources. They are described as follows.

1. SentiWordNet [4]: assigns three scores to Synsets of WordNet: positive score, negative score and objective score. When a word is looked up, the label corresponding to maximum of the three scores is returned. For multiple synsets of a word, the output label returned by majority of the Synsets becomes the prediction of the resource.
2. Subjectivity lexicon [18] is a resource that annotates words with tags like parts-of-speech, prior polarity, magnitude of prior polarity (weak/strong), etc. The prior polarity can be positive, negative or neutral. For prediction using this resource, we use this prior polarity.
3. General Inquirer [15] is a list of words marked as positive, negative and neutral. We use these labels to use Inquirer resource for our prediction.
4. Taboada [16] is a word-list that gives a count of collocations with positive and negative seed words. A word closer to a positive seed word is predicted to be positive and vice versa.

2.2 Feature Extraction

From this dataset we extracted a wide range of different features. The sentences were first parsed with Stanford POS Tagger, an English language parser (Kristina Toutanova and Christopher D. Manning. 2000.), which allowed us to extract linguistic information such as word tokens, lemmas, part-of-speech tags, syntactic functions and dependency structures. The actual feature vectors were then generated on the basis of this linguistic information by using a "bag of n-grams" approach, i.e. by constructing n-grams (unigrams, bigrams and trigrams) of each feature type (e.g. n-grams of word tokens, n-grams of part-of-speech tags...) and by counting for each n-gram in the training data how many times it occurs in the current instance. Additionally to these n-gram counts, we also included punctuation counts, average word length and average sentence length.

Sentiment Score Feature. Based on predictions of individual traits, we compute the Sentiment prediction for each trait with respect to a keyword in form of percentage of positive, negative and objective content. This is on the basis of predictions by each resource by weighting them according to their accuracies. These weights have been assigned to each resource based on experimental results. For each resource, the following scores are determined.

$$PositiveScore(s) = \sum_{i=0}^n P_i W_{P_i} \quad (1)$$

$$NegativeScore(s) = \sum_{i=0}^n N_i W_{N_i} \quad (2)$$

Table 2. LIWC features [13]

STANDARD COUNTS:
-Word count (WC), words per sentence (WPS), type/token ratio (Unique), words captured (Dic), words longer than 6 letters (Sixltr), negations (Negate), assents (Assent), articles (Article), prepositions (Preps), numbers (Number)
-Pronouns (Pronoun): 1st person singular (I), 1st person plural (We), total 1st person (Self), total 2nd person (You), total 3rd person (Other) PSYCHOLOGICAL PROCESSES:
-Affective or emotional processes (Affect): positive emotions (Posemo), positive feelings (Posfeel), optimism and energy (Optim), negative emotions (Negemo), anxiety or fear (Anx), anger (Anger), sadness (Sad)
-Cognitive Processes (Cogmech): causation (Cause), insight (Insight), discrepancy (Discrep), inhibition (Inhib), tentative (Tentat), certainty (Certain)
-Sensory and perceptual processes (Senses): seeing (See), hearing (Hear), feeling (Feel)
-Social processes (Social): communication (Comm), other references to people (Otherf), friends (Friends), family (Family), humans (Humans)
RELATIVITY:
-Time (Time), past tense verb (Past), present tense verb (Present), future tense verb (Future)
-Space (Space): up (Up), down (Down), inclusive (Incl), exclusive (Excl), Motion (Motion)
PERSONAL CONCERNS:
-Occupation (Occup): school (School), work and job (Job), achievement (Achieve)
-Leisure activity (Leisure): home (Home), sports (Sports), television and movies (TV), music (Music)
-Money and financial issues (Money)
-Metaphysical issues (Metaph): religion (Relig), death (Death), physical states and functions (Physcal), body states and symptoms (Body), sexuality (Sexual), eating and drinking (Eating), sleeping (Sleep), grooming (Groom)
OTHER DIMENSIONS:
-Punctuation (Allpct): period (Period), comma (Comma), colon (Colon), semi-colon (Semic), question (Qmark), exclamation (Exclam), dash (Dash), quote (Quote), apostrophe (Apostro), parenthesis (Parenth), other (Otherp)
-Swear words (Swear), nonfluencies (Nonfl), fillers (Fillers)

$$ObjectiveScore(s) = \sum_{i=0}^n O_i W_{O_i} \quad where, \quad (3)$$

$PositiveScore(s)$ = Positive score for each excerpts s ; $NegativeScore(s)$ = Negative score for each excerpts s ; $ObjectiveScore(s)$ = Objective score for each excerpts s ; n = Number of resources used for prediction; P_i, N_i, O_i = Positive, Negative, and Objective count of excerpt predicted respectively using resource i ; $W_{P_i}, W_{N_i}, W_{O_i}$ = Weights for respective classes derived for each resource i .

LIWC Features. We can extract features derived from the LIWC output. In specific, LIWC counts and groups the number of instances of nearly 4,500

keywords into 80 psychologically meaningful dimensions. We create one feature for each of the 80 LIWC dimensions, LIWC, 80 dimensions summarized mostly under the following four categories:

- Linguistic processes: Functional aspects of text (e.g., the average number of words per sentence, the rate of misspelling, swearing, etc.)
- Psychological processes: Includes all social, emotional, cognitive, perceptual and biological processes, as well as anything related to time or space.
- Personal concerns: Any references to work, leisure, money, religion, etc.
- Spoken categories: Primarily filler and agreement words.

For each instance, we calculate the ratio of words in each category from the LIWC toolkit [13], as these features are correlated with the personality dimensions [13]. These features and their categories are shown in Table 2.

2.3 Automated Approaches to Personality Classification

We explain three automated approaches to classify detecting personality behavior, each of which utilizes classifiers trained on the dataset of Section 2.1. The features employed by each strategy are described here.

Psycholinguistic Personality Detection. The Linguistic Inquiry and Word Count (LIWC) software [13] is a popular automated text analysis tool used widely in the social sciences. It has been used to detect personality traits [10], to study tutoring dynamics [2], and, most relevantly, to analyze personality detection [10].

Since LIWC software does not include a text classifier, we create features derived from the LIWC output. In particular, LIWC counts and groups the number of instances of nearly 4,500 keywords into 80 psychologically meaningful dimensions. We construct one feature for each of the 80 LIWC dimensions, which can be summarized broadly under the four categories that explained in Section 2.2. Indeed, the LIWC2007 software used in our experiments subsumes most of the features introduced in other work. Thus, we focus our psycholinguistic approach to personality detection on LIWC-based features.

2.4 Classification

On the other hand, our classification approach to personality detection provides us to model both content and context with n-gram features. Specifically, we consider the following two n-gram feature sets, with the corresponding features lowercased and unstemmed: UNIGRAMS and BIGRAMS. Features from the our approaches just introduced are used to train Naïve Bayes, Support Vector Machine classifiers, and Decision Tree.

Naïve Bayes (NB) Classifier. Naïve Bayes (NB) classifier provides a simple approach and can view such a classifier as a specialized form of Bayesian network and it leans on two simple assumptions. First, it assumes that the predictive attributes are conditionally independent given the class. Then, it posits that no hidden or latent attributes influence the prediction process [6].

For a document X , with label class c , the Naïve Bayes (NB) classifier gives us the following decision rule [6]:

$$P(C = c|X = x) = \frac{p(C = c)p(X = x|C = c)}{p(X = x)}, \text{ where} \quad (4)$$

$$P(X = x|C = c) = \prod_i P(X_i = x_i|C = c) \quad (5)$$

We use John and Langley [6] Naïve Bayes classifier in Weka [5] to train our Naive Bayes models on all three approaches and feature sets described above, namely LIWC, lexicons, UNIGRAMS, BIGRAMS. We also evaluate every combination of these features, but for brevity include only UNIGRAMS+BIGRAMS, which performs best.

Support Vector Machine (SVM). We also train Support Vector Machine (SVM) classifiers, which find a high-dimensional separating hyperplane between two groups of data. To simplify feature analysis in Section 5, we restrict our evaluation to linear SVMs, which learn a weight vector w and bias term b , such that a document x can be classified by:

$$y = \text{sign}(\vec{w} \cdot \vec{x} + b) \quad (6)$$

We use SMO [7] to train our SVM models on all three approaches and feature sets described above: LIWC, LEXICONS, UNIGRAMS, and BIGRAMS. We also evaluate every combination of these features, but for shortness include only LIWC+BIGRAMS, and LEXICON+BIGRAMS which performs best.

Decision Trees. We use J48, an open source Java implementation of the C4.5 algorithm in Weka [5] data mining tool to train our dataset for decision trees classifier. We evaluate all our approach on all combination of feature set, but we consider the features which performed best (UNIGRAMS+BIGRAMS, UNIGRAMS+LIWC). Our classification experiments are carried out with 10-fold cross validation on the corresponding dataset.

3 Results and Discussion

The model for classification personality strategies explained in Section 2 are performed using a 10-fold cross validation method under its default setting in Weka [5]. The parameters for model are chosen for each test fold based on standard cross validation experiments on the training dataset. All folds are chosen

Table 3. Automated classifier performance for three approaches based on 10-fold cross-validation experiments. Reported: Accuracy, Precision, Recall and F-measure are computed using Weka [5].

Approach	Features	Acc.	COM			DEP			LEA			WIT			COP			AGG		
			P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F
LEXICAL	lexicons _{j48}	61.95%	.67	.66	.67	.56	.66	.61	.61	.55	.58	.56	.54	.55	.66	.60	.63	.25	.21	.22
LIWC	liwc _{j48}	59.30%	.57	.62	.60	.64	.67	.65	.52	.40	.45	.62	.42	.50	.60	.62	.61	.50	.53	.51
CLASSIFIERS	unigrams _{svm}	60.54%	.74	.70	.72	.64	.63	.63	.50	.32	.50	.83	.39	.53	.52	.74	.61	.17	.33	.22
	bigrams _{svm}	70.40%	.92	.65	.76	.92	.75	.82	.79	.40	.52	1	.44	.62	.50	1	.66	1	.17	.29
	liwc+bigrams _{svm}	77.47%	.90	.75	.82	.93	.78	.85	.95	.64	.77	.93	.54	.68	.54	.95	.69	1	.2	.33
	lexicons+bigrams _{svm}	83.71%	.96	.80	.87	.96	.84	.90	.98	.76	.86	1	.74	.85	.98	.76	.86	1	.32	.48
	bigrams _{nb}	65.02%	.04	.87	.62	.87	.70	.77	.50	.21	.3	.80	.44	.57	.46	.96	.62	1	.16	.28
	unigrams+bigrams _{nb}	60.53%	.77	.60	.67	.72	.64	.68	.50	.53	.51	1	.39	.54	.40	.74	.52	.67	.5	.57
	unigrams+bigrams _{j48}	62.78%	.83	.67	.74	.83	.62	.71	.46	.43	.45	.82	.47	.60	.42	.84	.56	.26	.22	.24
	unigrams+liwc _{j48}	74.0%	.86	.80	.83	.81	.73	.77	.63	.64	.63	.85	.78	.81	.63	.75	.69	.50	.68	.57

so that each includes all instances from six classes; therefore, learned classifiers are always measured on dataset from unseen instances.

Table 3 shows the results of the top scores that we managed to achieve with each of the three classifiers over three approaches. We also use the combination of features and learner parameters that was determined to give the best accuracy by the classifiers. “Approach” column shows the model that have been tested, the “features” column indicates the types of features that have been used, the rest of columns indicates the results based on Accuracy, Precision, Recall, and F-measure (Acc., P, R, F) for all six classes.

We observe that our automated classifiers approaches achieve human judges (*kappa*) and baseline for most of feature sets. The statistical baseline for this six classes classification problem, considering the slight imbalances in the class distribution, is 30%. However there is an exception such as *Recall* for “aggressive” where does not performs significant. We can argue on this due to low number of instances in this class. However, this is expected given that human judges often focus on unreliable cues to aggressive utterances. If we look at the confusion matrix in Figure 3; Firstly, we note that most of the aggressive instances (8) classified as “helping” personality. Many other classes considered as “helping” as well. We figured out, this happened due to human judges evaluation, because the judges considered many small responses such as: OK, Yep, Thanks, Cool, etc as “helping” class. Secondly, as it shown in Table 1 the number of instances in “aggressive” class is low. We found out that the players are not often aggressive during chat conversation. It might be due to their work environment in that they are supervised by a human mentor during the game.

Interestingly, the psycholinguistic approach (LIWC_{j48}) performs almost 30% more accurately than baseline rather than SVM or NB. Also j48 perform higher than SVM and NB on lexical subjective scores features. In overall, all the standard text categorization approach proposed in Section 2 performs between 9% and 53% more accurately than baseline. However, best performance overall is achieved by combining features from these two approaches. Particularly, the combined model LEXICONS+BIGRAMS_{svm} is 83.71% accurate at personality classification.

Surprisingly, models trained only on UNIGRAMS_{svm} (60.54%), the simplest n-gram feature set, outperform LIWC (non text classification) approaches, and

a	b	c	d	e	f	<-- classified as
130	1	2	1	37	0	a = competitive
2	155	1	0	47	0	b = dependent
5	4	61	0	26	0	c = leading
0	0	0	14	12	0	d = withdrawn
2	1	0	0	146	0	e = helping
0	0	0	0	8	2	f = aggressive

Fig. 3. The Confusion Matrix performed by SVM Classifiers approach over BIGRAMS and subjective Lexicon features

Table 4. Top 16 highest weighted features learned by BIGRAMS+LEXICONS_{svm} and LIWC_{svm}. The results show for binary classification of “Helping, Aggressive” and “Leading, Dependent”.

BIGRAMS+LEXICONS _{svm}	LIWC _{svm}
Helping, Aggressive	Leading, Dependent
always want	six letters
didnt seem	pronoun
don t	personal pronoun
for me	i
is quite	we
it is	you
need to	she/he
no need	they
people don	impersonal pronouns
quite deadly	article
really that	verb
seem to	auxiliary verbs
slow down	past tense
speaking spanish	present tense
stop speaking	future tense

models trained on BIGRAMS_{nb} (65.02%) perform even better. This suggests that a universal set of feature such psycholinguistic keyword personality (i.e., LIWC) can not be the best model for personality detection, and a context-sensitive approach (e.g., BIGRAMS) might be necessary to achieve state-of-the-art personality detection performance.

To better understand the models learned by these automated approaches, we report in Table 4 the top 16 highest weighted features for two pair classes (Helping, Aggressive & Leading, Dependent) as learned by BIGRAMS+LEXICONS_{svm} and LIWC_{svm}. From BIGRAMS+LEXICONS_{svm} approach we have chosen classifier for classes “Helping” (with highest F-measure) and “Aggressive” (lowest F-measure), for LIWC_{svm} approach we have chosen classifier for classes “Leading, Dependent” with similar reason. We note that player with “Helping” personality behavior tend to use some how similar language with “Aggressive” players; in particular, “need to” and “no need”, the former one can

be consider as “Helping” behavior and latest one can be regarded as “Aggressive” attitude. Accordingly, in term of global features such as psycholinguistic features (LIWC), “Leading” and “Dependent” players tend to use similar pronouns(personal or impersonal) (i.e.; i, we, you, she/he, they). Finally, when we look at Confusion Matrix (Figure 3), it turns out that all misclassified instances from “Aggressive” class fall into “Helping” class and similarly almost 75% of misclassified instances in “Leading” class are classified as “Dependent” class.

4 Related Work

To our knowledge, the only research has been done specifically on the automatic classification of sentences based on Leary’s Rose for emotion detection is done by [17]. They described a methodology for a serious gaming project, deLearyous, which aims at developing an environment in which users can improve their communication skills by interacting with a virtual character in (Dutch) written natural language. In order to apply this framework, they classified the input sentences into one of four possible “emotion” classes (above, below, opp, tog, see Figure 1). They applied several machine learning algorithms, SVM, Naïve Bayes, Conditional Random field to obtained the classification performance. For this, they used different features set from the their dataset (unigrams, lemma trigrams and dependency structures). They obtained 52.5% accuracy around 25% over the baseline. In contrast, in our method we use Leary’s Rose framework to detect personality rather than emotion.

Mairesse et al [10,11] found that identification of personality (main Five in speech) by automatic analysis perform better than the baseline, and their analysis confirms previous findings linking language and personality, while revealing many new linguistic and prosodic markers. However, they had limitation for their method involving speech recognition that is recognition errors will introduce noise in all features except prosodic features, and prosodic features on their own are only effective in the extroversion model.

Another possible research that were done to let a machine learner determine the appropriate sentiment/emotion class. [12] and [8], for instance, attempt to classify LiveJournal posts according to their mood using Support Vector Machines trained with frequency features (word counts, POS-counts), length-related features (length of posts/sentences/...), semantic orientation features (using WordNet to calculate the distance of each word) and special symbols (emoticons).

5 Conclusion and Future Research

In this paper we have developed a dataset containing personality excerpts based on Leary’s Rose framework. By this, we have presented that the detection of personality behavior is more efficient than that of human judges. Consequently, we have presented three automated methods to personality detection, based on understanding from research in natural language processing, machine learning,

and psychology. We explore that while text classification based on n-gram (UNIGRAMS, BIGRAMS) is the best particular detection approach, a combination method such as LIWC and Subjective Lexicons features along with n-gram features can achieve better performance.

Eventually, we have done several notable contributions. Particularly, our results indicate it is vital to take into account both the context, such as BIGRAMS, rather than precisely using a global set of personality indications (e.g., LIWC and Subjective Lexicons). We have also shown some findings based on the feature weights found by our classifiers that show the difficulties confronted by judges in annotating the dataset. Finally, we have found a possible connection between personality behavior by players, such “Helping, Aggressive” and “Dependent, Leading”, based on BIGRAMSs and LIWC similarities.

For future work, we want to include an extended experiment of the methods proposed in current research to sentiment analysis, opinion mining, as well as emotion detection in other domains. Also, we want to extend the method in this work to apply in Big-Five personality detection. It will help us to not only detect the player’s behaviors but also to detect introvert and extrovert players, and a focus on approaches with POS features might be useful.

Acknowledgements. This work was funded by the National Science Foundation (DRK-12-0918409). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of these funding agencies, cooperating institutions, or other individuals

References

1. Bagley, E.A.S.: Stop Talking and Type: Mentoring in a Virtual and Face-to-Face Environmental Education Environment. Ph.D. thesis, University of Wisconsin-Madison (2011)
2. Cade, W.L., Lehman, B.A., Olney, A.: An exploration of off topic conversation. In: Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, pp. 669–672. Association for Computational Linguistics, Los Angeles (June 2010), <http://www.aclweb.org/anthology/N10-1096>
3. D’Angelo, C., Arastoopour, G., Chesler, N., Shaer, D.: Collaborating in a virtual engineering internship. In: Computer Supported Collaborative Learning Conference (CSCL), Hong Kong, SAR (2011)
4. Esuli, A., Sebastian, F.: Sentiword-net: A publicly available lexical resource for opinion mining. In: Proceedings of LREC 2006, Genova, Italy (2006)
5. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA data mining software: An update. SIGKDD Explorations 11(1) (2009)
6. John, G.H., Langley, P.: Estimating continuous distributions in bayesian classifiers. In: Eleventh Conference on Uncertainty in Artificial Intelligence, San Mateo, pp. 338–345 (1995)
7. Keerthi, S., Shevade, S., Bhattacharyya, C., Murthy, K.: Improvements to platt’s smo algorithm for svm classifier design. Neural Computation 13(3), 637–649 (2001)

8. Keshtkar, F., Inkpen, D.: Using sentiment orientation features for mood classification in blogs, pp. 1–6 (September 2009)
9. Leary, T.: *The Interpersonal Diagnosis of Personality*. John Wiley and Sons Inc. (1957)
10. Mairesse, F., Walker, M., Mehl, M., Moore, R.: Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research* 30(1), 457–500 (2007)
11. Mairesse, F., Walker, M.: Automatic recognition of personality in conversation. In: *Proceedings of the Human Language Technology Conference of the North American Chapter of the ACL*, New York, pp. 85–88 (2006)
12. Mishne, G.: Experiments with mood classification in blog posts. *ACM SIGIR* (2005)
13. Pennebaker, J.W., Chung, C.K., Ireland, M., Gonzales, A., Booth, R.J.: *Linguistic inquiry and word count (liwc)*. Lawrence Erlbaum Associates, Mahwah (2007)
14. Shaer, D.W., D’Angelo, C., Chesler, N.C., Arastoopour, G.: *Nephrotex: Teaching first year students how to think like engineers*. In: *Laboratory Improvement (CCLI) PI Conference*, Washington D.C. (2011)
15. Stone, P., Dunphy, D., Smith, M., Ogilvie, D.: *The general inquirer: A computer approach to content analysis*. MIT Press (1966)
16. Taboada, M., Grieve, J.: Analyzing appraisal automatically. In: *Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications*, Stanford, CA, pp. 158–161 (2004)
17. Vaassen, F., Daelemans, W.: Emotion classification in a serious game for training communication skills. In: *Proceedings of the 20th Meeting of Computational Linguistics in the Netherlands*, Netherlands, pp. 155–168 (2010)
18. Wiebe, J., Wilson, T.: Learning to disambiguate potentially subjective expressions. In: *Proceedings of the Conference on Natural Language Learning (CoNLL)*, pp. 112–118 (2002)